

Informatica**U**manistica

Introduzione alle Biblioteche Digitali



UNIVERSITÀ DI PISA

Sommario [1/2]

◆ Cenni storici

- Vannevar Bush
- Dalle Biblioteche ai Cataloghi Automatizzati
- Gli OPAC accessibili via Web
- Le Biblioteche Digitali

◆ Cos'è una Biblioteca Digitale

- Definizione
- Confronto tra BD e database, sistemi IR, WWW, biblioteca tradizionale
- Vantaggi delle BD
- Alcuni esempi di Biblioteche Digitali

Sommario [2/2]

- ◆ **Cosa ha permesso la nascita delle Biblioteche Digitali**
 - Evoluzioni tecnologiche

- ◆ **Tipologie di Biblioteche Digitali**
 - Biblioteche Pubbliche e Biblioteche Specializzate
 - Tipi di documenti trattati
 - Libri
 - Documenti testuali
 - Immagini
 - Audio/video
 -

Cenni storici

Vannevar Bush (1890-1974)

- ◆ **Direttore dell'US Office of Scientific Research and Development**
- ◆ **Ha predetto diverse evoluzioni tecnologiche**
 - L'idea del "MEMEX" (1945) conteneva molte delle idee sulle quali si basano il Web e le Biblioteche Digitali
 - Il MEMEX aveva lo scopo di fornire ai ricercatori la possibilità di scambiarsi facilmente informazioni e di avere accesso alla totalità delle conoscenze disponibili

Memex

- ◆ **Integrazione di computer, tastiera e scrivania**
- ◆ **“archivio meccanizzato privato e biblioteca”**
 - Eliminazione degli aspetti che rendevano ripetitivo e inefficace il retrieval di informazione
 - Implementazione basata sull'uso di microfilm
- ◆ **Indicizzazione associativa**
 - “il processo di legare due elementi insieme è l'aspetto più importante”
 - Preludio agli ipertesti...

Memex

- ◆ **L'informazione poteva essere indicizzata utilizzando associazioni tra i vari elementi, considerando anche l'adattamento alle esigenze degli utenti**
 - WWW non lo fornisce ancora oggi
- ◆ **Bush osservava che i nuovi strumenti modificano il modo in cui facciamo le cose ed espandono le nostre esigenze**
 - L'impatto del WWW e delle Biblioteche Digitali non è ancora completamente noto
- ◆ **L'articolo di Bush non prevedeva l'uso di sistemi di ricerca free-text**
 - knowledge trails only; Yahoo minus keyword searching

Dalle Biblioteche tradizionali alle Biblioteche Digitali

◆ Biblioteche tradizionali

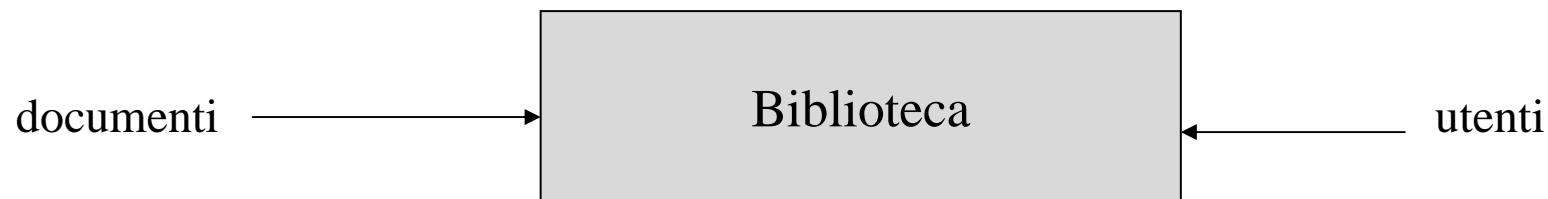
- Inizialmente erano luoghi dedicati alla conservazione dei documenti
- Con l'aumento della quantità di documenti immagazzinati, si è reso necessario creare degli strumenti di ricerca efficaci
- L'avvento dei calcolatori ha permesso di automatizzare e rendere più efficienti ed efficaci gli strumenti di ricerca

◆ Biblioteche digitali

- I documenti stessi sono in forma digitale
 - ➔ Ricerca basata sul contenuto
 - ➔ Conservazione degli oggetti digitali
 - ➔ Protezione da accessi indesiderati, ecc.

Le Biblioteche tradizionali

- ◆ La biblioteca svolge il ruolo di “mediatore” tra gli oggetti “portatori di informazione” (documenti) e gli “utilizzatori dell’informazione”



- ◆ **DOCUMENTO:** Qualsiasi oggetto utilizzabile a fini di consultazione, ricerca, informazione

Funzioni delle biblioteche

◆ SELEZIONE

- In una biblioteca è necessario selezionare il materiale da conservare, sia per quanto riguarda il formato (ad es. libri, filmati, foto, ecc.) che il contenuto.

◆ ACQUISIZIONE

- Dopo aver individuato cosa inserire o conservare nella biblioteca, è necessario acquisire fisicamente l'opera.

◆ DESCRIZIONE

- Una volta inserita la nuova opera nella biblioteca, questa deve essere descritta sia per quanto riguarda il contenuto che per quanto riguarda aspetti quali la conservazione, i diritti di accesso, ecc.

◆ ACCESSO

- La biblioteca fornisce degli strumenti per la ricerca e l'accesso delle opere conservate nella biblioteca.

◆ CONSERVAZIONE

- La biblioteca è inoltre responsabile di preservare le opere presenti. Aspetti che devono essere considerati riguardano la conservazione fisica delle opere, la loro protezione da danni esterni, ecc.

Strumenti per la ricerca

- ◆ Anticamente la ricerca avveniva sulla base della posizione fisica dei documenti. In base a questa procedura le opere venivano poste in determinati scaffali sulla base del nome dell'autore o della categoria alla quale appartenevano.
- ◆ Questa procedura viene seguita ancora oggi per l'organizzazione delle biblioteche che contengono opere non digitali (ad es. libri)
- ◆ Ma attualmente si utilizzano dei **cataloghi elettronici** che contengono una descrizione delle opere e permettono di effettuare ricerche sia sulla base degli autori che di altri attributi, quali ad esempio la casa editrice, parole chiave che descrivono l'opera, ecc.

Strumenti per la ricerca

- ◆ I documenti vengono ordinati sulla base di un singolo criterio di posizionamento fisico all'interno della biblioteca
- ◆ Ad esempio, i documenti sono collocati sulla base della
 - ➔ Data di “arrivo” (liste inventariali)
 - ➔ Soggetto (classificazione)
 - ➔ Tipo di documento

...e possono essere ricercati soltanto in base a quel criterio

Strumenti per la ricerca

◆ Il catalogo:

- Ciascun documento è descritto con un insieme di elementi significativi, scelti secondo le regole della descrizione bibliografica (Titolo, autori, ... soggetto, ... collocazione); la descrizione è riportata su una o più schede
 - Ciascuna scheda è intestata con uno degli elementi bibliografici ritenuti utili a ricercare il documento ==> Punti di accesso
 - Le schede intestate sono ordinate nel catalogo secondo i punti di accesso
-

◆ *Il catalogo permette la ricerca di un documento conoscendo uno qualsiasi dei suoi punti di accesso*

La varietà delle -teche

[depositi==>collezioni, raccolte]

- ◆ **Biblioteche** per i libri
- ◆ **Emeroteche** per le opere periodiche
- ◆ **Cineteche** per i filmati
- ◆ **Discoteche** per i dischi
- ◆ **Mediateche** per la conservazione di opere di tipo diverso
- ◆ **Archivi**
- ◆ **Musei** per le opere d'arte

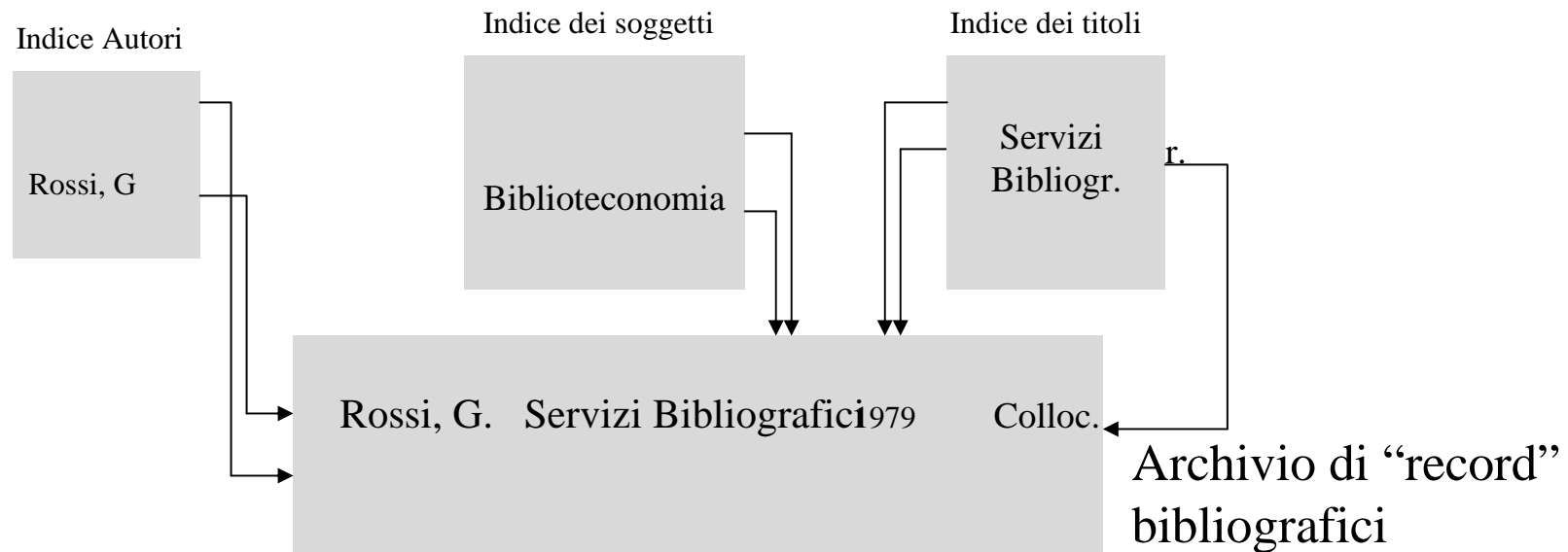
La varietà delle -teche

- ◆ Ciascuna “-teca” raccoglie e organizza una speciale tipologia di documenti
- ◆ Ciascuna tipologia di documenti ha proprie regole di descrizione

.....ma rimangono valide le regole di organizzazione dei cataloghi

I cataloghi “automatizzati”

- ◆ L’uso dei calcolatori ha reso le prestazioni dei cataloghi più potenti e più flessibili
- ◆ In particolare possono essere facilmente gestiti degli indici per semplificare e velocizzare la ricerca, la selezione e l’accesso ai documenti
 - ◆ **Indici** : liste ordinate di elementi particolari della descrizione bibliografica, utili a identificare il documento di interesse : Autori - titoli - soggetti - collane -



I cataloghi “automatizzati”

- ◆ Nei cataloghi a schede si può ricercare un documento solo attraverso uno dei suoi punti di accesso
- ◆ Nei cataloghi automatizzati, invece, si possono fare interrogazioni definendo più punti di accesso:

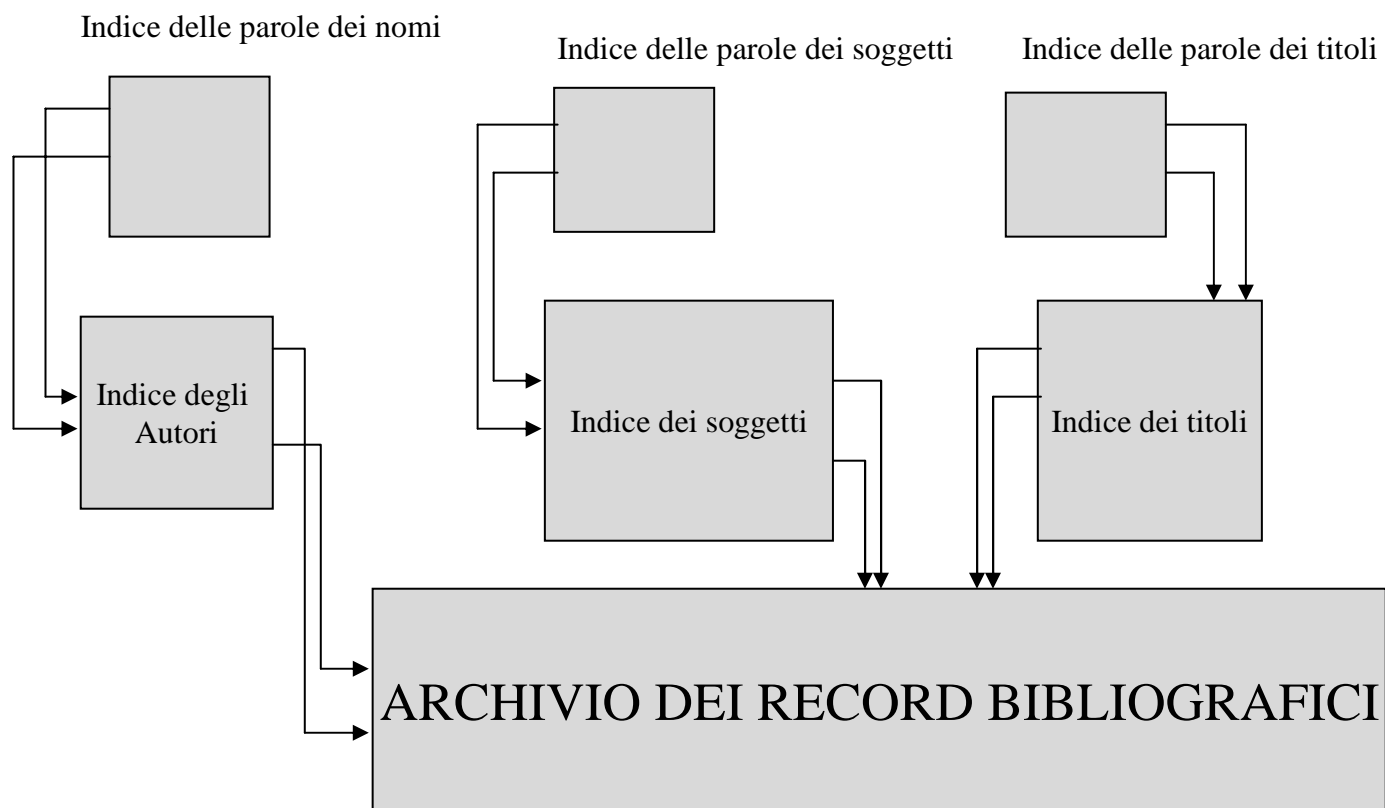
AUT = Rossi, G.

AND

TIT = Servizi bibliografici

I cataloghi “automatizzati”

I cataloghi di ultima generazione costruiscono “indici degli indici”



I cataloghi “automatizzati”

Ricerche con dati non certi:

AUT = Rossi AND TIT = Servizi

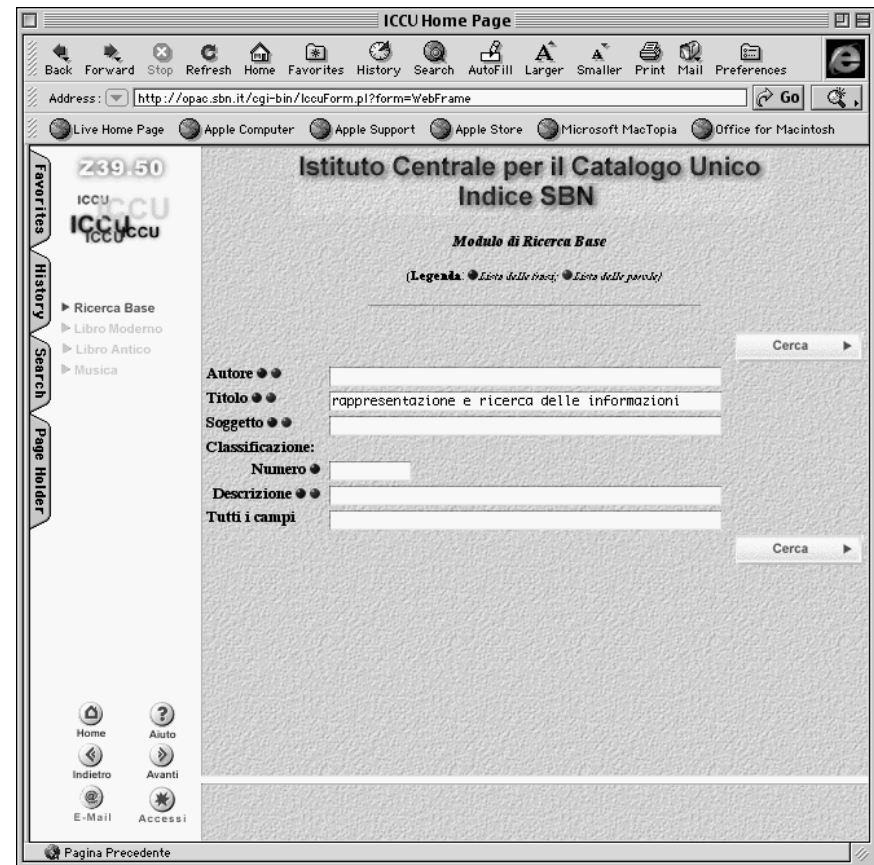
TIT = bibliogr#

PAROLA = biblio# AND DATA > 1999

Il catalogo delle biblioteche automatizzate

- ◆ Un'ulteriore caratteristica delle biblioteche automatizzate, riguarda l'uso di una *interfaccia utente-sistema*
 - Tale interfaccia è il modo in cui il sistema informativo si presenta all'utente per istruirlo e guidarlo a svolgere le operazioni
 - Ogni biblioteca digitale avrà la propria interfaccia, che dipenderà anche dalle funzionalità che vengono offerte all'utente dal sistema

Il catalogo del Servizio Bibliografico Nazionale: WWW.sbn.it



Cos'è una Biblioteca Digitale

Definizione informale

Una Biblioteca Digitale è una collezione organizzata di oggetti digitali accessibili in rete, ed un insieme di servizi che permettono la conservazione, l'accesso e la ricerca, oltre che l'organizzazione e la manutenzione della collezione. La collezione può contenere dati di tipo diverso, quali ad esempio testi, immagini, audio, video, ecc. il cui contenuto è descritto sulla base di un opportuno modello di metadati.

Gli oggetti digitali

- ◆ **Una Biblioteca Digitale permette di archiviare “documenti” contenenti vari tipi di dati**
 - Testo
 - Immagini
 - Video
 - Audio
 - 3D objects
 - Virtual-reality worlds
 -
 - Composizione dei tipi di dati precedenti

I principali servizi di una Biblioteca Digitale

◆ Accesso e recupero

- Cataloghi
- Riferimenti
- Indici

◆ Conservazione

◆ Gestione

- Controllo dell'accesso
- Condivisione dei dati
- Gestione della collaborazione tra gli utenti
 - E.g. collaborative filtering, catalogazione,
-

◆ Una Biblioteca Digitale dovrà fornire almeno tutti I servizi offerti da una Biblioteca tradizionale

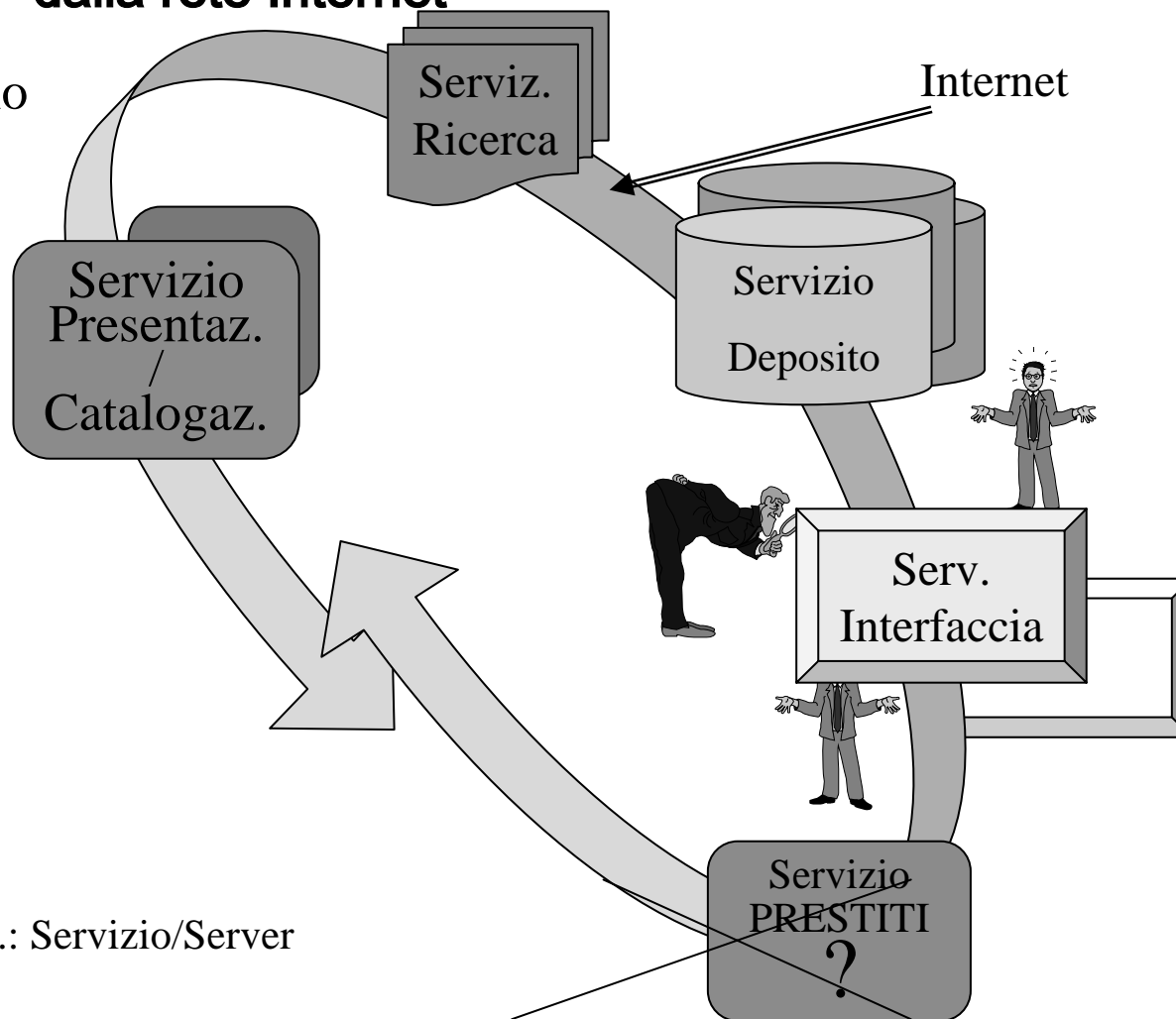
Che cos'è una Biblioteca Digitale?

Un sistema distribuito per la gestione di risorse digitali accessibili dalla rete Internet

Gli utenti accedono al servizio
Interfaccia WEB
attraverso il
Protocollo HTTP

La comunicazione fra i
servizi
del sistema avviene
attraverso
la rete, con l'uso di
un protocollo
specifico del sistema di DL

N.B.: Servizio/Server



In cosa differisce una Biblioteca Digitale da un DataBase

◆ Un database tradizionale SQL ha la relazione come contenitore base dell'informazione:

Trovare i nomi degli impiegati che lavorano in progetti con budget superiore a 200000 Euro

```
select Nome
from Impegati, Progetti
where Impegati.NumProj = Progetti.NumProj AND
      Progetti.Budget > 200000
```

◆ I database sfruttano strutture e relazioni note

◆ Il retrieval dei DBMS non è probabilistico

Impiegati

| Nome | NumProj | Age | Salario |
|---------|---------|-----|---------|
| Rossi | 25 | 39 | 39000 |
| Bianchi | 27 | 45 | 43000 |
| Verdi | 25 | 29 | 30000 |

Progetti

| NumProj | Titolo | Budget |
|---------|-----------------|--------|
| 25 | Biblioteche Dig | 220000 |
| 26 | Sviluppo sw | 190000 |
| 27 | Sviluppo hw | 180000 |

In cosa differisce una BD da un sistema di IR tradizionale?

- ◆ La differenza è meno netta che per i DBMS
- ◆ I sistemi IR possono essere considerati precursori delle BD
- ◆ I sistemi di IR si sono occupati tradizionalmente di documenti testuali ma loro evoluzioni trattano anche documenti multimediali
 - Match esatto - Boolean, text pattern searching
 - Match non esatto - probabilistic, vector space, clustering
- ◆ Le BD possono essere considerate un superset dei sistemi IR

In cosa differisce una BD dal WWW?

- ◆ **La differenza principale è legata all'organizzazione dei documenti**
 - Il WWW non impone alcuna organizzazione dell'informazione, mentre nelle BD tutte le operazioni sono soggette a procedure ben definite.
 - In particolare, nelle BD sono definiti precisi schemi di metadati che permettono l'accesso ai documenti
- ◆ **Alcuni motori di ricerca del web (Yahoo, Lycos) tentano di aggiungere una qualche organizzazione ai documenti trattati**
 - Comunque, non tutti i documenti del web sono gestiti
 - La maggior parte dei motori di ricerca si basa su text search (Altavista, Google)

In cosa differisce una BD dal WWW?

- ◆ **Un'altra differenza sostanziale è legata al controllo degli inserimenti**
 - I documenti nel web possono essere inseriti da chiunque, mentre in una DB l'inserimento è permesso solo a particolari utenti
 - I motori di ricerca selezionano i documenti da indicizzare tra quelli presenti nel web, mentre nelle BD tutti i documenti vengono indicizzati sulla base dei criteri definiti

- ◆ **Le DB sono soggette ad un maggior controllo (per gli inserimenti, gli accessi e le ricerche) del WWW, ed hanno un insieme di utenti ben preciso**

Quali sono le differenze tra una DB ed una Biblioteca Tradizionale (BT)?

◆ Le BT gestiscono oggetti (documenti) fisici

- Anche se le BT utilizzano delle schede elettroniche per individuare i documenti, questi si trovano in una ben precisa posizione fisica
- Questo porta a delle ovvie implicazioni
 - **Gli oggetti possono esistere solo in un luogo**
 - **Un solo utente per volta può accedere all'oggetto**
 - **L'oggetto può essere acceduto solo recandosi fisicamente presso la Biblioteca o attraverso meccanismi di distribuzione postale**

Quali sono le differenze tra una DB ed una Biblioteca Tradizionale (BT)?

◆ Biblioteca Digitale

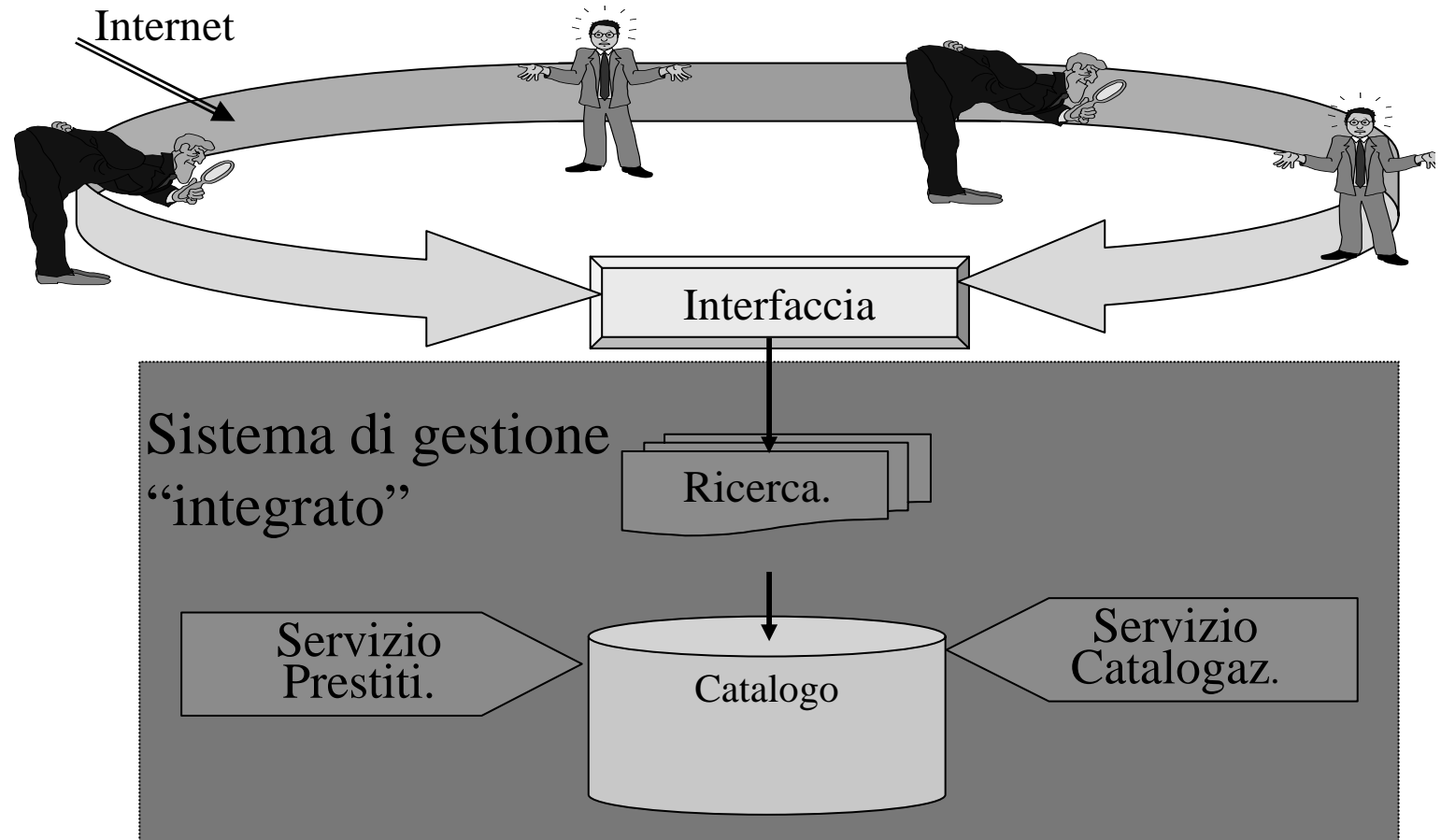
- Rimuovendo la restrizione della esistenza fisica degli oggetti si ottengono notevoli vantaggi
 - **Accessi multipli, trasmissione elettronica**
- ma anche alcune complicazioni
- **Proprietà intellettuale, diritti di accesso, etc.**

◆ Una BT offre anche vantaggi dal punto di vista sociale ed educativo

- Molte BT offrono servizi aggiuntivi che non possono essere offerti da una BD (ambiente di conversazione, comunicazione tra i lettori, ...) almeno per ora

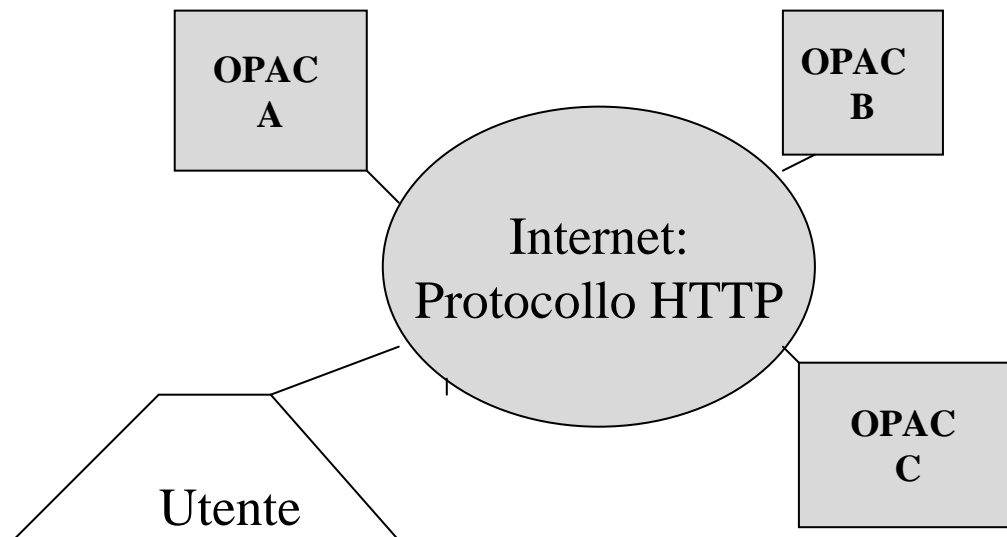
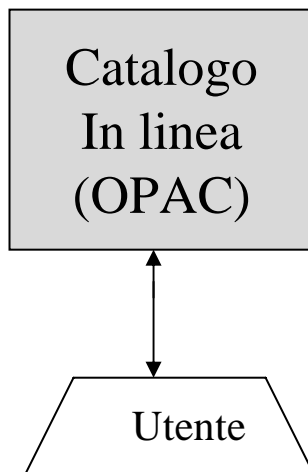
Cos'è una BIBLIOTECA DIGITALE ?

Analogie con un sistema di gestione per biblioteche "tradizionali"



Cos'è una BIBLIOTECA DIGITALE ?
Analogie con un sistema di gestione per biblioteche "tradizionali":
La comunicazione in rete: protocolli di comunicazione

Dalla comunicazione
"diretta"



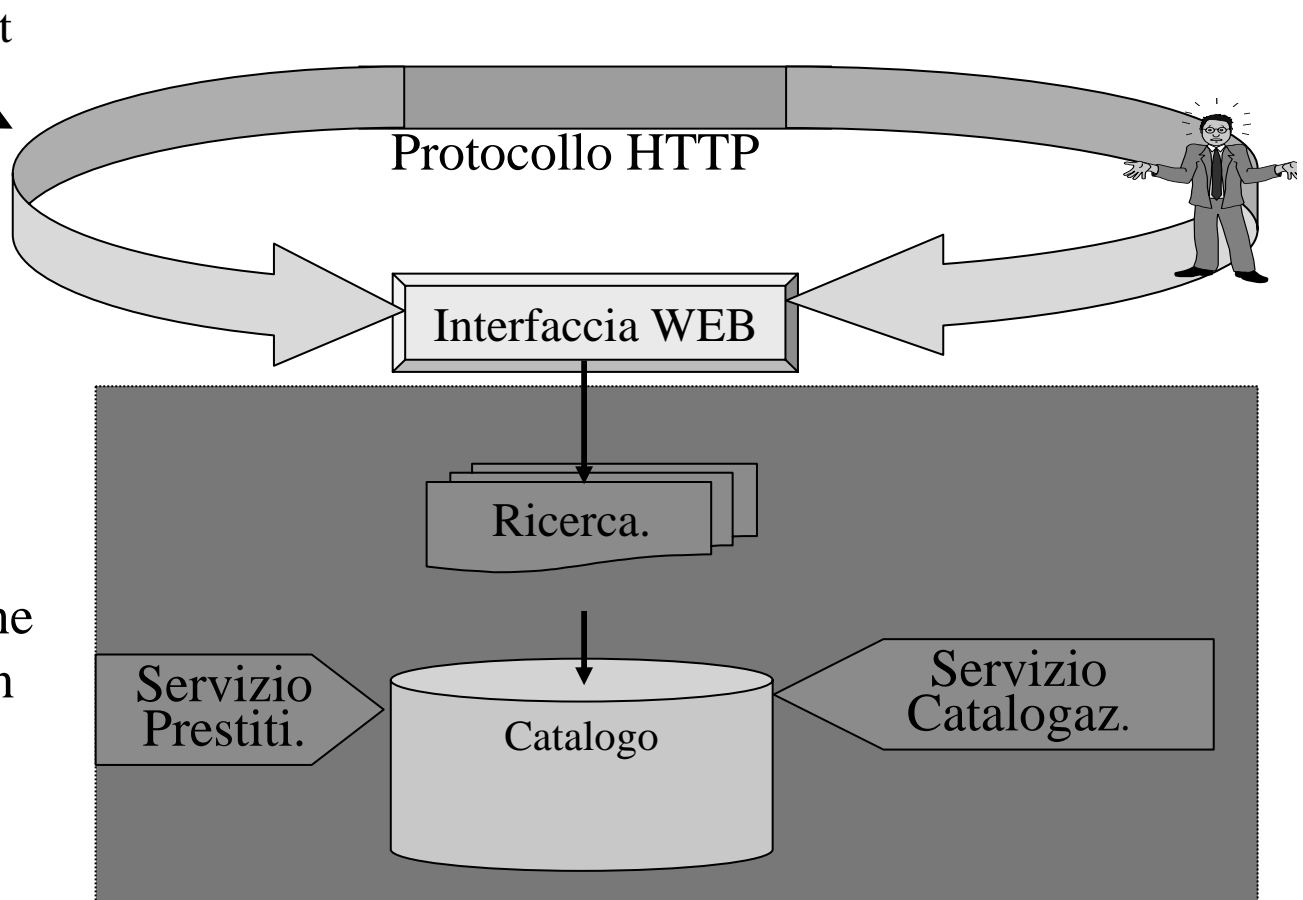
....alla comunicazione in
rete via WEB: protocollo
HTTP

Cos'è una BIBLIOTECA DIGITALE ?
Analogie con un sistema di gestione per biblioteche "tradizionali":
La comunicazione in rete fra utente e sistema:
protocolli di comunicazione

Internet

Gli utenti generici
accedono
via WEB soltanto ai
servizi di Informazione,
Ricerca/Recupero

I vari moduli software che
realizzano i "servizi" non
hanno esigenze di
comunicazione in rete



Cos'è una BIBLIOTECA DIGITALE ?

Analogie con un sistema di gestione per biblioteche “tradizionali”

I servizi “di base” di una biblioteca digitale:

- **Interfaccia**
- **Presentazione degli oggetti digitali (“Acquisto” e catalogazione)**
- **Deposito**
- **Ricerca/Browsing/Recupero degli oggetti digitali**

Vantaggi delle Biblioteche Digitali

◆ Vantaggi per gli utenti

- Si costruisce un BD nella convinzione di poter fornire una migliore distribuzione dell'informazione
 - **Informazione disponibile dove è necessario**
 - **Maggiori quantità di informazione disponibile**
 - **Possibilità di selezionare facilmente quello che interessa**
 - **Possibilità di utilizzare media diversi (testo, immagini, audio, video, ecc.)**
 - **L'informazione può essere condivisa**
 - **L'informazione è sempre aggiornata**
 - **Accesso 24/24**

Vantaggi delle BD

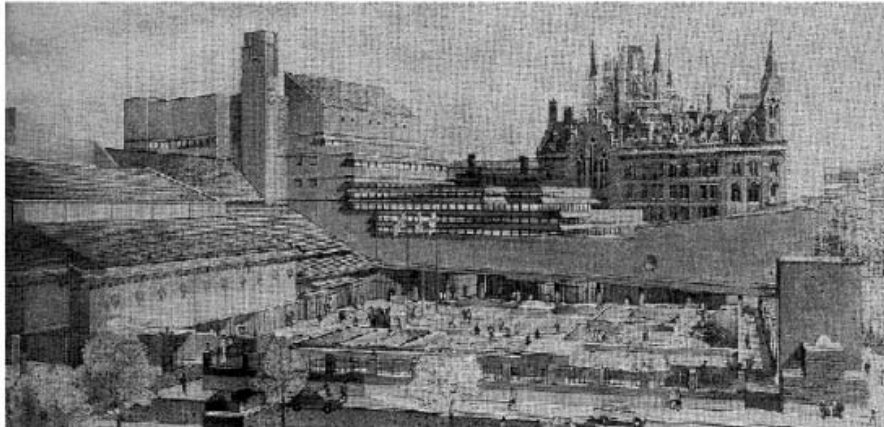
◆ Vantaggi economici

- Le biblioteche convenzionali risultano sempre più costose
 - Infrastrutture
 - Personale
 - Pubblicazione
- Attualmente anche le BD hanno costi elevati ma
 - I costi sono destinati a scendere, in particolare i costi di archiviazione e distribuzione

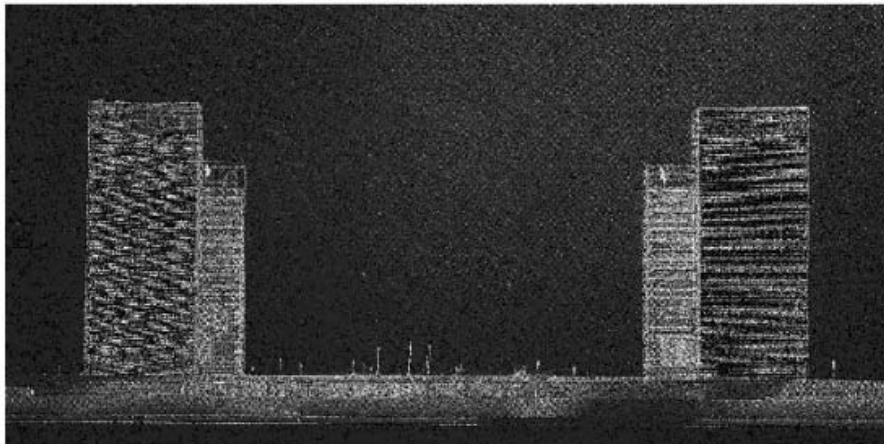
◆ Le BD richiedono la definizione di nuovi modelli dei costi



Each of these libraries cost more to build than the cost of scanning its books



The British Library, London £450M



The Bibliothèque de France, Paris FFr 5B

from Lesk,
<http://community.bellcore.com/lesk/columbia/session1/>

Alcuni esempi di Biblioteche Digitali

- ◆ **Esempi di cataloghi di Biblioteche Digitali (Chemical Abstracts, Inspec, Biblioteca del CNR di Pisa)**

- ◆ **Biblioteche Pubbliche**
 - Library of Congress
 - New York Public Library
 - Boston Public Library

- ◆ **Biblioteche Specializzate**
 - NTRSDL
 - Documentari storici (DB ECHO)

Cosa ha permesso la nascita delle Biblioteche Digitali

Un po' di storia

- ◆ **I primi tentativi di archiviare informazione bibliotecaria con i computer datano dalla fine degli anni '60. I problemi riguardavano**
 - Alto costo dei computer
 - Interfaccia Uomo/Macchina scadente
 - Mancanza di reti di comunicazione

- ◆ **Primi risultati**
 - Library of Congress ha memorizzato le schede MARC (formato per machine readable cataloguing)

- ◆ **Architettura dei sistemi**
 - Piccole quantità di informazione memorizzata su un computer centrale
 - Gli utenti erano connessi a terminali remoti di bassa qualità e con bassa velocità di connessione al computer centrale

Evolutioni tecnologiche [1/3]

- ◆ **Archiviazione elettronica sta diventando sempre più economica rispetto alla carta**
 - Gli edifici delle Biblioteche tradizionali impegnano circa $\frac{1}{4}$ dei costi
 - Le Biblioteche tradizionali hanno problemi di espansione (alti costi, tempi lunghi, ecc.)
 - I costi di archiviazione elettronica tendono a diminuire (circa 30% per anno)

Evolutioni tecnologiche [2/3]

◆ **Miglioramento della qualità dei display**

- Miglioramento della risoluzione
- Disponibilità di software di visualizzazione
- Disponibilità di standard di visualizzazione

◆ **Reti ad alta velocità**

- Connessione diffusa
- Aumento della velocità dei link della rete
- Aumento della velocità delle connessioni locali
- In alcuni paesi è più facile (e veloce) ricevere informazione tramite la rete internet che a stampa

Evolutioni tecnologiche [3/3]

◆ Accesso alla Biblioteca

- Le Biblioteche tradizionali sono accessibili solo agli utenti dell'organizzazione
 - **Esistono organizzazioni con biblioteche molto fornite (per es. centri medici specializzati) ma molti utenti non possono accedervi**
- L'accesso ad una Biblioteca Digitale richiede costi sempre più bassi
 - **Meno di 1000 Euro per il computer**
 - **Meno di 10 Euro/mese per la rete**

Esempi di cataloghi automatizzati

Esempi

- ◆ Library of Congress
- ◆ Boston Public Library
- ◆ British Library
- ◆ New York Public Library
- ◆ Chemical Abstracts
- ◆ Biblioteca del CNR di Pisa
- ◆ Meta OPAC di Pisa
- ◆ Il catalogo del Servizio Bibliografico Nazionale

Biblioteche Digitali Multimediali

- ◆ **Museo del Louvre**
 - Archivio delle opere
- ◆ **Uffizi**
- ◆ **Istituto Luce**
- ◆ **Le teche RAI**

Differenze tra una BD e ...

... una Biblioteca Tradizionale

- ◆ La biblioteca ETRDL
- ◆ E-books della Boston Public Library

... il WWW

- ◆ Ricerca di documenti con Google
- ◆ Il catalogo di Yahoo